

INTERNATIONAL COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Assistant Commissioner for Patents
United States Patent and Trademark
Office
Box PCT
Washington, D.C. 20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 06 December 1999 (06.12.99)	
International application No. PCT/DE99/01308	Applicant's or agent's file reference 98P1695P
International filing date (day/month/year) 03 May 1999 (03.05.99)	Priority date (day/month/year) 11 May 1998 (11.05.98)
Applicant HOLZAPFEL, Martin	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:

02 November 1999 (02.11.99)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was

☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Margret Fourné-Godbersen
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro



INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁶: G10L 5/04	A1	(11) Internationale Veröffentlichungsnummer: WO 99/59134 (43) Internationales Veröffentlichungsdatum: 18. November 1999 (18.11.99)
(21) Internationales Aktenzeichen: PCT/DE99/01308 (22) Internationales Anmeldedatum: 3. Mai 1999 (03.05.99) (30) Prioritätsdaten: 198 21 031.0 11. Mai 1998 (11.05.98) DE (71) Anmelder (für alle Bestimmungsstaaten ausser US): SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D-80333 München (DE). (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HOLZAPFEL, Martin [DE/DE]; Stabelerstrasse 13, D-80933 München (DE). (74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 17 34, D-80506 München (DE).		(81) Bestimmungsstaaten: JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Veröffentlicht <i>Mit internationalem Recherchenbericht. Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist; Veröffentlichung wird wiederholt falls Änderungen eintreffen.</i>
(54) Title: METHOD AND DEVICE FOR DETERMINING SPECTRAL VOICE CHARACTERISTICS IN A SPOKEN EXPRESSION (54) Bezeichnung: VERFAHREN UND ANORDNUNG ZUR BESTIMMUNG SPEKTRALER SPRACHCHARAKTERISTIKA IN EINER GESPROCHENEN ÄUSSERUNG (57) Abstract <p>According to the invention, spectral voice characteristics are determined in a natural language expression, whereby the expression is digitized and subjected to a wavelet transformation. The speaker-specific characteristics arise from the different transformation steps of the wavelet transformation. Within the scope of a voice synthesis, these characteristics can be compared with characteristics of other expressions in order to generate a continuously sounding synthetic voice signal for the human ear. Alternatively, the characteristics can also be modified in a targeted manner in order to counteract a perceptive dissonance.</p> (57) Zusammenfassung <p>Es werden spektrale Sprachcharakteristika in einer natürlichsprachlichen Äußerung bestimmt, wobei die Äußerung digitalisiert und einer Wavelet-Transformation unterzogen wird. Aus den unterschiedlichen Transformationsstufen der Wavelet-Transformation gehen die sprecherspezifischen Charakteristika hervor. Diese Charakteristika können im Rahmen einer Sprachsynthese mit Charakteristika anderer Äußerungen verglichen werden, um ein für das menschliche Ohr kontinuierlich klingendes synthetisches Sprachsignal zu erzeugen. Alternativ können die Charakteristika auch gezielt verändert werden, um einer perceptiven Dissonanz entgegenzuwirken.</p>		

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Beschreibung**Verfahren und Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung**

5

Die Erfindung betrifft ein Verfahren und eine Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung.

- 10 Bei einer konkatenativen Sprachsynthese werden einzelne Laute aus Sprachdatenbanken zusammengesetzt. Um dabei einen für das menschliche Ohr natürlich klingenden Sprachverlauf zu erhalten, sind Diskontinuitäten an den Punkten, wo die Laute zusammengesetzt werden (Konkatenationspunkte) zu vermeiden.
- 15 Die Laute sind dabei insbesondere Phoneme einer Sprache oder eine Zusammensetzung mehrerer Phoneme.

- Eine Wavelet-Transformation ist aus [1] bekannt. Bei der Wavelet-Transformation ist durch ein Wavelet-Filter
- 20 gewährleistet, daß jeweils ein Hochpaßanteil und ein Tiefpaßanteil einer nachfolgenden Transformationsstufe ein Signal einer aktuellen Transformationsstufe vollständig wiederherstellen. Dabei erfolgt von einer Transformationsstufe zur nächsten eine Reduktion der
- 25 Auflösung des Hochpaßanteils bzw. Tiefpaßanteils (engl. Fachbegriff: "Subsampling"). Insbesondere ist durch das Subsampling die Anzahl der Transformationsstufen endlich.

- Die **Aufgabe** der Erfindung besteht darin, ein Verfahren und
- 30 eine Anordnung zur Bestimmung spektraler Sprachcharakteristika anzugeben, mit deren Hilfe insbesondere eine natürlich wirkende synthetische Sprachausgabe bestimmbar ist.

- 35 Diese Aufgabe wird gemäß den Merkmalen der unabhängigen Patentansprüche gelöst.

Im Rahmen der Erfindung wird ein Verfahren angegeben zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung. Dazu wird die gesprochene Äußerung digitalisiert und einer Wavelet-Transformation unterzogen.

- 5 Anhand unterschiedlicher Transformationsstufen der Wavelet-Transformation werden die sprecherspezifischen Charakteristika ermittelt.

- 10 Dabei ist es insbesondere ein Vorteil, daß bei der Wavelet-Transformation mittels eines Hochpaßfilters und eines Tiefpaßfilters die Äußerung aufgeteilt wird und unterschiedliche Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen sprecherspezifische Charakteristika enthalten.

- 15 Die einzelnen Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen stehen für vorgegebene sprecherspezifische Charakteristika, wobei sowohl Hochpaßanteil als auch Tiefpaßanteil einer jeweiligen Transformationsstufe, also das jeweilige Charakteristikum, getrennt von anderen Charakteristika modifiziert werden kann. Setzt man bei der inversen Wavelet-Transformation aus den jeweiligen Hochpaß- und Tiefpaßanteilen der einzelnen Transformationsstufen wieder das ursprüngliche Signal zusammen, so ist gewährleistet, daß genau das gewünschte Charakteristikum verändert worden ist. Es ist somit möglich bestimmte vorgegebene Eigenarten der Äußerung zu verändern, ohne daß dadurch der Rest der Äußerung beeinflusst wird.

- 20 25 30 Eine Ausgestaltung besteht darin, daß vor der Wavelet-Transformation die Äußerung gefenstert, also eine vorgegebene Menge von Abtastwerten ausgeschnitten, und in den Frequenzbereich transformiert wird. Hierzu wird insbesondere eine Fast-Fourier-Transformation (FFT) angewandt.

- 35 Eine weitere Ausgestaltung besteht darin, daß ein Hochpaßanteil einer Transformationsstufe in einen Realteil

und einen Imaginärteil aufgeteilt wird. Der Hochpaßanteil der Wavelet-Transformation entspricht dem Differenzsignal zwischen dem aktuellen Tiefpaßanteil und dem Tiefpaßanteil der vorhergehenden Transformationsstufe.

5

Insbesondere besteht eine Weiterbildung darin, die Zahl der durchzuführenden Transformationsstufen der Wavelet-Transformation dadurch zu bestimmen, daß in der letzten Transformationsstufe, die aus hintereinandergeschalteten
10 Tiefpässen besteht, ein Gleichanteil der Äußerung enthalten ist. Dann ist das Signal als Ganzes darstellbar durch seine Wavelet-Koeffizienten. Dies entspricht der vollständigen Transformation der Information des Signalausschnitts in den Wavelet-Raum.

15

Wird insbesondere nur der jeweilige Tiefpaßanteil weiter transformiert (mittels eines Hochpaß- und eines Tiefpaßfilters), so verbleibt als Hochpaßanteil einer Transformationsstufe das Differenzsignal, wie oben erläutert.
20 Kumuliert man Differenzsignale (Hochpaßanteile) über die Transformationsstufen, erhält man in der letzten Transformationsstufe als kumulierten Hochpaßanteil die Information der gesprochenen Äußerung ohne Gleichanteil.

25 Im Rahmen einer zusätzlichen Weiterbildung sind die sprecherspezifischen Charakteristika identifizierbar als:

a) Grundfrequenz:

Die Schwingung des Hochpaßanteils der ersten oder der
30 zweiten Transformationsstufe der Wavelet-Transformation läßt die Grundfrequenz der Äußerung erkennen. Die Grundfrequenz zeigt an, ob der Sprecher ein Mann oder einen Frau ist.

35

b) Form der spektralen Hüllkurve:

Die spektrale Hüllkurve enthält Information über eine Transferfunktion des Vokaltrakts bei der Artikulation.

5 In einem stimmhaften Bereich wird die spektrale
Hüllkurve von den Formanten dominiert. Der
Hochpaßanteil einer höheren Transformationsstufe der
Wavelet-Transformation enthält diese spektrale
Hüllkurve.

c) Spectral Tilt (Rauchigkeit):

10 Die Rauchigkeit in einer Stimme wird als negative
Steigung im Verlauf des vorletzten Tiefpaßanteils
sichtbar.

Die sprecherspezifischen Charakteristika a) bis c) sind bei
der Sprachsynthese von großer Bedeutung. Wie eingangs
erwähnt, bedient man sich bei der konkatenativen
15 Sprachsynthese großer Mengen realgesprochener Äußerungen, aus
denen Beispiellaute ausgeschnitten und später zu einem neuen
Wort zusammengesetzt werden (synthetisierte Sprache). Dabei
sind Diskontinuitäten zwischen zusammengesetzten Lauten von
Nachteil, da diese vom menschlichen Ohr als unnatürlich
20 wahrgenommen werden. Um den Diskontinuitäten entgegenzuwirken
ist es von Vorteil, direkt die perzeptiv relevanten Größen zu
erfassen und ggf. zu vergleiche und/oder einander anzupassen.

Dies kann geschehen durch direkte Manipulation, indem ein
25 Sprachlaut in mindestens einer seiner sprecherspezifischen
Charakteristika angepaßt wird, so daß er in dem akustischen
Kontext der konkatenativ verknüpften Laute nicht als störend
wahrgenommen wird. Auch ist es möglich, die Auswahl eines
passenden Lautes daran auszurichten, daß sprecherspezifische
30 Charakteristika von zu verknüpfenden Lauten möglichst gut
zueinander passen, z.B. daß den Lauten gleiche oder ähnliche
Rauchigkeit zu eigen ist.

Ein Vorteil der Erfindung besteht darin, daß die spektrale
35 Hüllkurve den Artikulationstrakt des Sprechers widerspiegelt
und nicht, wie z.B. ein Polstellenmodell, auf Formanten
gestützt ist. Weiterhin gehen bei der Wavelet-Transformation

als nichtparametrischer Darstellung keine Daten verloren, die Äußerung kann stets vollständig rekonstruiert werden. Die aus den einzelnen Transformationsstufen der Wavelet-Transformation hervorgehenden Daten sind linear voneinander unabhängig, können somit getrennt voneinander beeinflusst und später wieder zu der beeinflussten Äußerung - verlustlos - zusammengesetzt werden.

- 10 Weiterhin wird eine Anordnung zur Bestimmung spektraler Sprachcharakteristika angegeben, die eine Prozessoreinheit aufweist, die derart eingerichtet ist, daß eine Äußerung digitalisierbar ist. Daraufhin wird die Äußerung einer Wavelet-Transformation unterzogen und anhand
- 15 unterschiedlicher Transformationsstufen werden sprecherspezifische Charakteristika ermittelt.

Diese Anordnung ist insbesondere geeignet zur Durchführung des erfindungsgemäßen Verfahrens oder einer seiner vorstehend

20 erläuterten Weiterbildungen.

Weiterbildungen der Erfindung ergeben sich auch aus den abhängigen Ansprüchen.

- 25 Ausführungsbeispiele der Erfindung werden nachfolgend anhand der Zeichnung dargestellt und erläutert.

Es zeigen

- 30 Fig.1 eine Wavelet-Funktion;
- Fig.2 eine Wavelet-Funktion, unterteilt nach Realteil und Imaginärteil;
- 35 Fig.3 eine kaskadierte Filterstruktur, die die Transformationsschritte der Wavelet-Transformation darstellt;

Fig.4 Tiefpaßanteile und Hochpaßanteile unterschiedlicher Transformationsstufen;

5 Fig.5 Schritte der konkatenativen Sprachsynthese.

Fig.1 zeigt eine Wavelet-Funktion, die bestimmt ist durch

$$10 \quad \psi(f) = c \cdot \left(1 - \left(\frac{f}{\sigma}\right)^2\right) \cdot e^{-\frac{1}{2} \cdot \left(\frac{f}{\sigma}\right)^2} \quad (1),$$

wobei

f die Frequenz,

σ eine Standardabweichung und

15 c eine vorgegebene Normierungskonstante bezeichnen.

Insbesondere ist die Standardabweichung σ bestimmt durch die vorgebbare Stelle des Seitenbandminimums 101 in Fig.1.

20

Fig.2 zeigt eine Wavelet-Funktion mit einem Realteil gemäß Gleichung (1) und einer Hilbert-Transformierten H des Realteils als Imaginärteil. Die komplexe Wavelet-Funktion ergibt sich somit zu

25

$$\Psi(f) = \psi(f) + j \cdot H\{\psi(f)\} \quad (2).$$

Die Konstante c aus Gleichung (1) wird verwendet, um die komplexe Wavelet-Funktion zu normieren:

30

$$\int_{-\infty}^{\infty} \Psi(f) \cdot \bar{\Psi}(f) \, df = 1 \quad (3),$$

wobei $\bar{\Psi}$ die konjugiert komplexe Wavelet-Funktion bezeichnet.

Fig.3 zeigt die kaskadierte Anwendung der Wavelet-Transformation. Ein Signal 301 wird sowohl durch einen Hochpaß HP1 302 als auch durch einen Tiefpaß TP1 305 gefiltert. Dabei findet insbesondere ein Subsampling statt, d.h. die Anzahl der abzuspeichernden Werte wird pro Filter reduziert. Eine inverse Wavelet-Transformation gewährleistet, daß aus dem Tiefpaßanteil TP1 305 und dem Hochpaßanteil HP1 304 wieder das ursprüngliche Signal 301 rekonstruierbar ist.

Im Hochpaß HP1 302 wird getrennt nach Realteil Re1 303 und Imaginärteil Im1 304 gefiltert.

Das Signal 310 nach dem Tiefpaßfilter TP1 305 wird erneut sowohl durch einen Hochpaß HP2 306 als auch durch einen Tiefpaß TP2 309 gefiltert. Der Hochpaß HP2 306 umfaßt wieder einen Realteil Re2 307 und einen Imaginärteil Im2 308. Das Signal nach der zweiten Transformationsstufe 311 wird wieder gefiltert, usf.

Geht man von einem (FFT-transformierten) Kurzzeitspektrum mit 256 Werten aus, so werden acht Transformationsschritte durchgeführt (Subsamplingrate: 1/2), bis das Signal aus dem letzten Tiefpaßfilter TP8 dem Gleichanteil entspricht.

In Fig.4 sind verschiedene Transformationsstufen der Wavelet-Transformation, unterteilt nach Tiefpaßanteilen (Figuren 4A, 4C und 4E) und Hochpaßanteilen (Figuren 4B, 4D und 4F) dargestellt.

Aus dem Hochpaßanteil gemäß Fig.4B ist die Grundfrequenz der gesprochenen Äußerung ersichtlich. Neben den Schwankungen in der Amplitude ist deutlich eine überwiegende Periodizität im wavelet-gefilterten Spektrum zu erkennen, die Grundfrequenz des Sprechers. Anhand der Grundfrequenz ist es möglich, vorgegebene Äußerungen bei der Sprachsynthese einander

anzupassen oder passende Äußerungen aus einer Datenbank mit vorgegebene Äußerungen zu bestimmen.

Im Tiefpaßanteil von Fig.4C sind als ausgeprägte Minima und
5 Maxima die Formanten des Sprachsignalausschnitts (die Länge des Sprachsignalausschnitts entspricht in etwa der doppelten Grundfrequenz) dargestellt. Die Formanten repräsentieren Resonanzfrequenzen im Vokaltrakt des Sprechers. Die deutliche Darstellbarkeit der Formanten ermöglicht eine Anpassung
10 und/oder Auswahl passender Lautbausteine bei der konkatenativen Sprachsynthese.

Im Tiefpaßanteil der vorletzten Transformationsstufe (bei 256 Frequenzwerten im Originalsignal: TP7), kann die Rauchigkeit
15 einer Stimme ermittelt werden. Der Abstieg des Kurvenverlaufs zwischen Maximum Mx und Minimum Mi kennzeichnet den Grad der Rauchigkeit.

Die erwähnten drei sprecherspezifischen Charakteristika sind
20 somit identifiziert und können für die Sprachsynthese gezielt beeinflußt werden. Dabei ist es insbesondere von Bedeutung, daß bei der inversen Wavelet-Transformation die Manipulation eines einzelnen sprecherspezifischen Charakteristikums nur dieses beeinflußt, die anderen perzeptiv relevanten Größen
25 bleiben unberührt. Somit kann die Grundfrequenz gezielt verstellt werden, ohne daß dadurch die Rauchigkeit der Stimme beeinflußt wird.

Eine andere Einsatzmöglichkeit besteht in der Auswahl eines
30 geeigneten Lautabschnitts zur konkatenativen Verknüpfung mit einem anderen Lautabschnitt, wobei beide Lautabschnitte ursprünglich von verschiedenen Sprechern in unterschiedlichen Kontexten aufgenommen wurden. Mit Ermittlung spektraler Sprachcharakteristika kann ein geeigneter zu verknüpfender
35 Lautabschnitt gefunden werden, da mit den Charakteristika Kriterien bekannt sind, die einen Vergleich von Lautabschnitten untereinander und somit eine Auswahl des

passenden Lautabschnitts automatisch nach bestimmten Vorgaben ermöglichen.

Fig.5 zeigt Schritte einer konkatenativen Sprachsynthese.

- 5 Eine Datenbank wird mit einer vorgegebenen Menge natürlichgesprochener Sprache verschiedener Sprecher erstellt, wobei Lautabschnitte in der natürlichgesprochenen Sprache identifiziert und abgespeichert werden. Es ergeben sich zahlreiche Repräsentanten für die verschiedenen
- 10 Lautabschnitte einer Sprache, auf die die Datenbank zugreifen kann. Die Lautabschnitte sind insbesondere Phoneme einer Sprache oder eine Aneinanderreihung solcher Phoneme. Je kleiner der Lautabschnitt, desto größer sind die Möglichkeiten bei der Zusammensetzung neuer Wörter. So umfaßt
- 15 die deutsche Sprache eine vorgegebene Menge von ca. 40 Phonemen, die zur Synthese nahezu aller Wörter der Sprache ausreichen. Dabei sind unterschiedliche akustische Kontexte zu berücksichtigen, je nachdem, in welchem Wort das jeweilige Phonem auftritt. Nun ist es wichtig, die einzelnen Phoneme in
- 20 den akustischen Kontext derart einzubetten, daß Diskontinuitäten, die vom menschlichen Gehör als unnatürlich und "synthetisch" empfunden werden, vermieden werden. Wie erwähnt stammen die Lautabschnitte von unterschiedlichen Sprechern und weisen somit verschiedene sprecherspezifische
- 25 Charakteristika auf. Um eine möglichst natürlich wirkende Äußerung zu synthetisieren, ist es wichtig, die Diskontinuitäten zu minimieren. Dies kann erfolgen durch Anpassung der identifizierbaren und modifizierbaren sprecherspezifischen Charakteristika oder durch Auswahl
- 30 passender Lautabschnitte aus der Datenbank, wobei ebenfalls die sprecherspezifischen Charakteristika bei der Auswahl ein entscheidendes Hilfsmittel darstellen.

- In Fig.5 sind beispielhaft zwei Laute A 507 und B 508
- 35 dargestellt, die jeweils einzelne Lautabschnitte 505 bzw. 506 aufweisen. Die Laute A 507 und B 508 stammen jeweils aus einer gesprochenen Äußerung, wobei der Laut A 507 deutlich

10

vom Laut B 508 verschieden ist. Eine Trennlinie 509 zeigt an, wo der Laut A 507 mit dem Laut B 508 verknüpft werden soll. Im vorliegenden Fall sollen die ersten drei Lautabschnitte des Lautes A 507 mit den letzten drei Lautabschnitten des
5 Lautes B 508 konkatenativ verknüpft werden.

Es wird entlang der Trennlinie 509 ein zeitliches Dehnen oder Stauchen (vergleiche Pfeil 503) der aufeinanderfolgenden Lautabschnitte durchgeführt, um den diskontinuierlichen
10 Eindruck am Übergang 509 zu vermindern.

Eine Variante besteht in einem abrupten Übergang der entlang der Trennlinie 509 geteilten Laute. Dabei kommt es jedoch zu den erwähnten Diskontinuitäten, die das menschliche Gehör als
15 störend wahrnimmt. Fügt man hingegen einen Laut C zusammen, daß die Lautabschnitte innerhalb eines Übergangsbereichs 501 oder 502 berücksichtigt werden, wobei ein spektrales Abstandsmaß zwischen zwei einander zuordenbaren
Lautabschnitten in dem jeweiligen Übergangsbereich 501 oder
20 502 angepaßt wird (allmählicher Übergang zwischen den Lautabschnitten). Als das Abstandsmaß herangezogen wird insbesondere im Wavelet-Raum der euklidische Abstand zwischen den in diesem Bereich relevanten Koeffizienten.

25

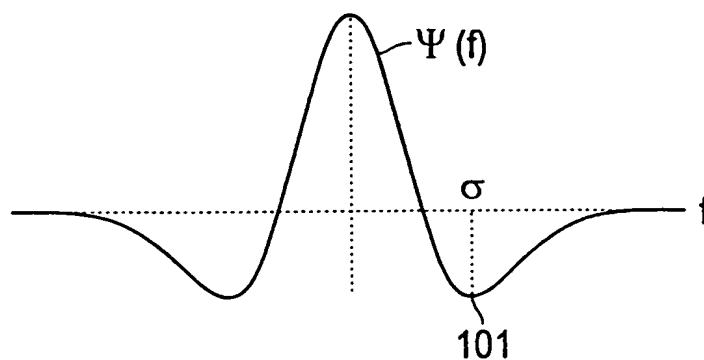
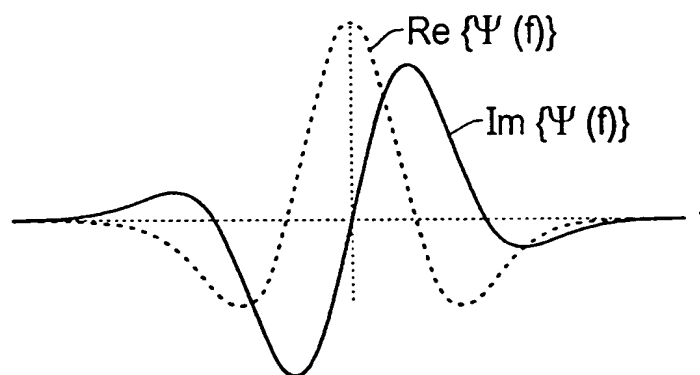
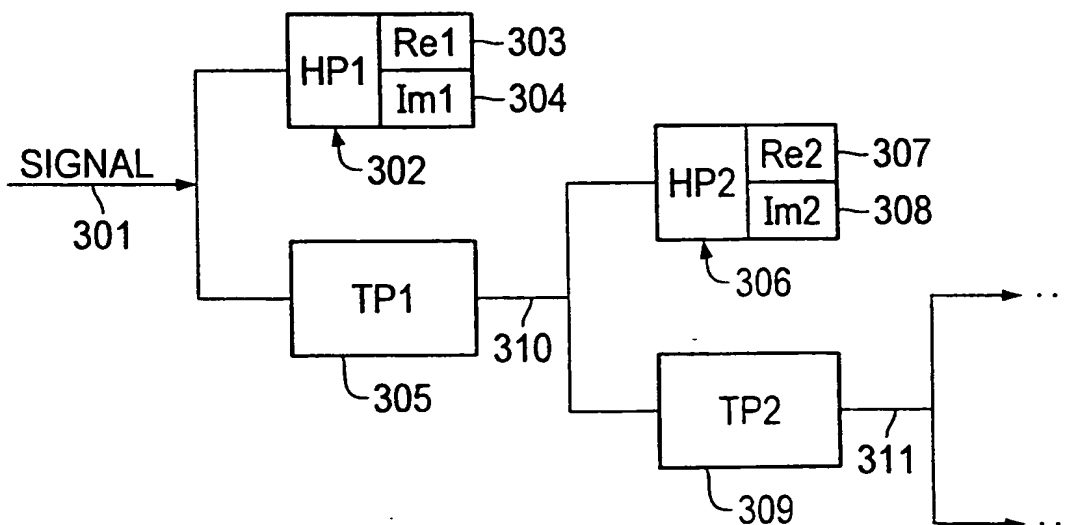
Literaturverzeichnis:

- [1] I. Daubechies: "Ten Lectures on Wavelets", Siam Verlag
1992, ISBN 0-89871-274-2, Kapitel 5.1, Seiten 129-137.

Patentansprüche

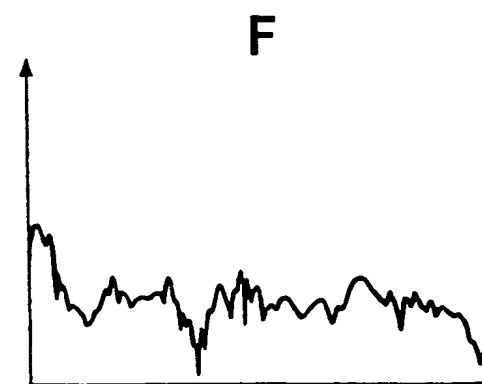
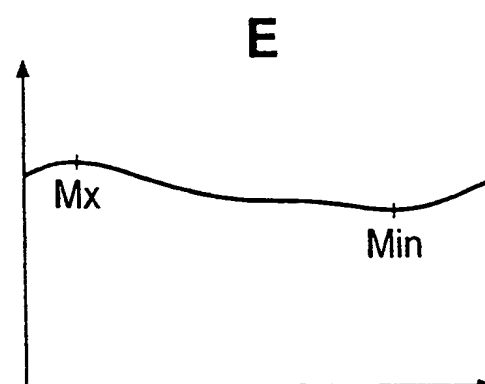
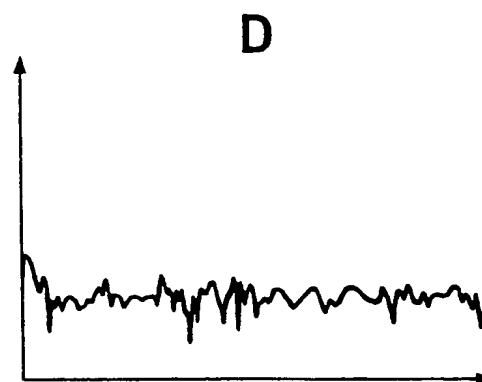
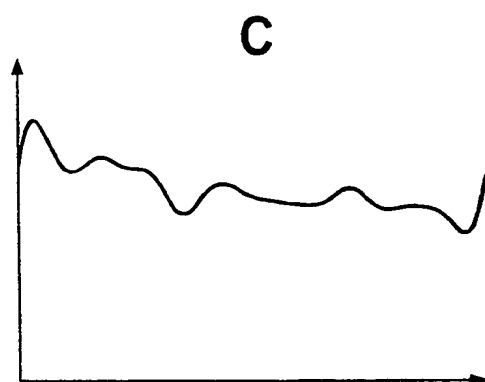
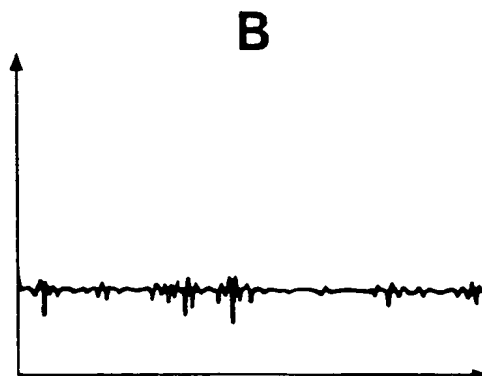
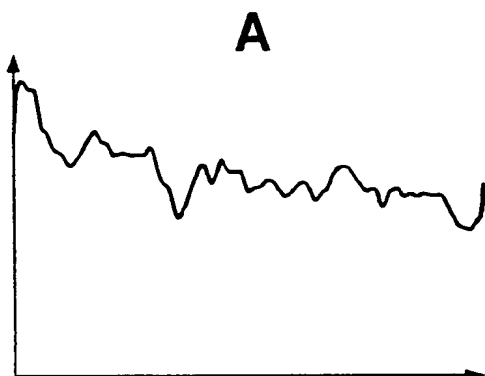
1. Verfahren zur Bestimmung spektraler Sprachcharakteristika
in einer gesprochenen Äußerung,
 - 5 a) bei dem die Äußerung digitalisiert wird,
 - b) bei dem die digitalisierte Äußerung einer Wavelet-Transformation unterzogen wird,
 - c) bei dem anhand unterschiedlicher Transformationsstufen
der Wavelet-Transformation die sprecherspezifischen
10 Charakteristika bestimmt werden.
2. Verfahren nach Anspruch 1,
bei dem vor der Wavelet-Transformation eine gefensterte
Transformation der digitalisierten Äußerung in einen
15 Frequenzbereich durchgeführt wird.
3. Verfahren nach Anspruch 2,
bei dem die Transformation in den Frequenzbereich mittels
Fast-Fourier-Transformation durchgeführt wird.
20
4. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem in jeder Stufe der Wavelet-Transformation ein
Tiefpaßanteil und ein Hochpaßanteil eines zu
transformierenden Signals ermittelt werden.
25
5. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem ein Hochpaßanteil nach einem Realteil und einem
Imaginärteil unterteilt wird.
- 30 6. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die Wavelet-Transformation mehrere
Transformationsstufen umfaßt, wobei die letzte
Transformationsstufe einen Gleichanteil der Äußerung in
einer der Anzahl Transformationsstufen entsprechenden
35 wiederholten Tiefpaßfilterung liefert.

7. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die sprecherspezifischen Charakteristika bestimmt
sind durch:
a) eine Grundfrequenz der gesprochenen Äußerung;
5 b) spektrale Hüllkurve;
c) einer Rauchigkeit der gesprochenen Äußerung.
8. Verwendung des Verfahrens nach einem der Ansprüche 1 bis
7 zur Sprachsynthese,
10 wobei einzelne sprecherspezifische Charakteristika im
Hinblick auf eine natürlich klingende Aneinanderreihung
von Sprachlauten angepaßt werden.
9. Verwendung des Verfahrens nach einem der Ansprüche 1 bis
15 7 zur Sprachsynthese,
wobei aus einer vorgegebenen Datenmenge diejenigen
Sprachlaute anhand einzelner spektraler
Sprachcharakteristika ausgewählt werden, die eine
natürlich klingende Aneinanderreihung von Sprachlauten
20 gewährleisten.
10. Anordnung zur Bestimmung spektraler Sprachcharakteristika
in einer gesprochenen Äußerung
25 mit eineressoreinheit, die derart eingerichtet ist,
daß folgende Schritte durchführbar sind:
a) die Äußerung wird digitalisiert;
b) die digitalisierte Äußerung wird einer Wavelet-
Transformation unterzogen;
30 c) anhand unterschiedlicher Transformationsstufen der
Wavelet-Transformation werden die sprecherspezifischen
Charakteristika bestimmt.

1/3
FIG 1**FIG 2****FIG 3**

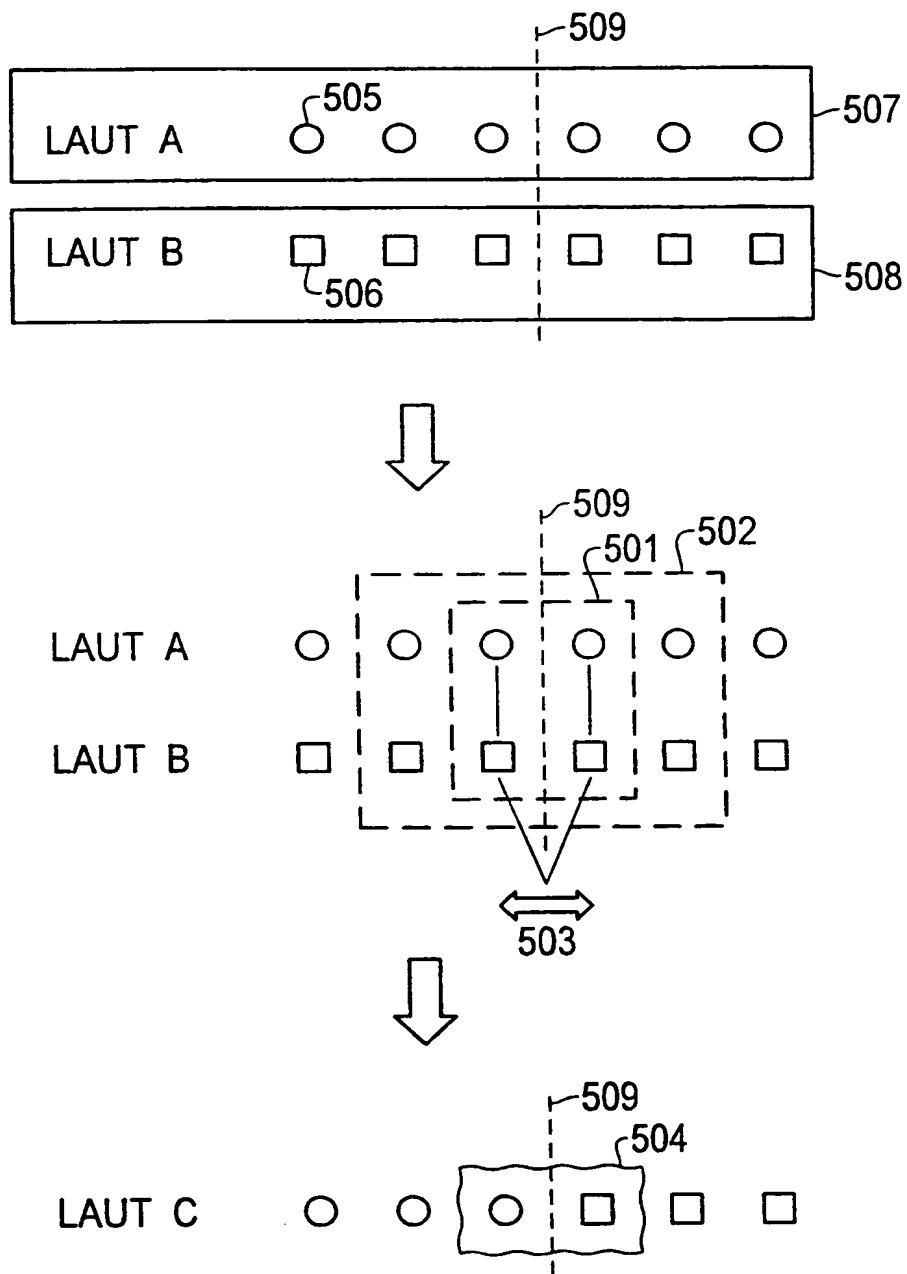
2/3

FIG 4



3/3

FIG 5



INTERNATIONAL SEARCH REPORT

International Application No

PC1/DE 99/01308

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G10L5/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 528 725 A (HUI SIEW K) 18 June 1996 (1996-06-18) column 5, line 9 - line 24; figure 2 column 6, line 57 - column 7, line 47; figure 5	1-4, 6, 7, 10
Y	---	8, 9
Y	EP 0 519 802 A (SEXTANT AVIONIQUE) 23 December 1992 (1992-12-23) page 3, column 3, line 57 - page 4, line 26; figures 3-6	8, 9
A	EP 0 703 565 A (IBM) 27 March 1996 (1996-03-27) abstract page 4, line 5 - page 6, line 34; figures 2, 3 --- -/--	1, 8-10

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

5 October 1999

Date of mailing of the international search report

12/10/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Wanzeele, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 99/01308

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EVANGELISTA G: "PITCH-SYNCHRONOUS WAVELET REPRESENTATIONS OF SPEECH AND MUSIC SIGNALS"</p> <p>IEEE TRANSACTIONS ON SIGNAL PROCESSING, vol. 41, no. 12, 1 December 1993 (1993-12-01), pages 3313-3330, XP000426651</p> <p>ISSN: 1053-587X</p> <p>abstract</p> <p>paragraph '00II!; figure 1</p> <p>paragraph '000V!</p> <p>paragraph '0V.A!; figure 7</p> <p>-----</p>	1,10

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/DE 99/01308

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5528725	A	18-06-1996	GB 2272554 A	18-05-1994
EP 0519802	A	23-12-1992	FR 2678103 A	24-12-1992
			WO 9222890 A	23-12-1992
			JP 6503186 T	07-04-1994
			US 5826232 A	20-10-1998
EP 0703565	A	27-03-1996	JP 8095589 A	12-04-1996
			US 5671330 A	23-09-1997

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT VON DEM GEBIET DES PATENTWESENS

PCT

INTERNATIONALER RECHERCHENBERICHT

(Artikel 18 sowie Regeln 43 und 44 PCT)

Aktenzeichen des Anmelders oder Anwalts 98P1695P	WEITERES VORGEHEN siehe Mitteilung über die Übermittlung des internationalen Recherchenberichts (Formblatt PCT/ISA/220) sowie, soweit zutreffend, nachstehender Punkt 5	
Internationales Aktenzeichen PCT/DE 99/ 01308	Internationales Anmeldedatum (Tag/Monat/Jahr) 03/05/1999	(Frühestes) Prioritätsdatum (Tag/Monat/Jahr) 11/05/1998
Anmelder SIEMENS AKTIENGESELLSCHAFT et al.		

Dieser internationale Recherchenbericht wurde von der Internationalen Recherchenbehörde erstellt und wird dem Anmelder gemäß Artikel 18 übermittelt. Eine Kopie wird dem Internationalen Büro übermittelt.

Dieser internationale Recherchenbericht umfaßt insgesamt 3 Blätter.

☒ Darüber hinaus liegt ihm jeweils eine Kopie der in diesem Bericht genannten Unterlagen zum Stand der Technik bei.

1. Grundlage des Berichts

a. Hinsichtlich der **Sprache** ist die internationale Recherche auf der Grundlage der internationalen Anmeldung in der Sprache durchgeführt worden, in der sie eingereicht wurde, sofern unter diesem Punkt nichts anderes angegeben ist.

☐ Die internationale Recherche ist auf der Grundlage einer bei der Behörde eingereichten Übersetzung der internationalen Anmeldung (Regel 23.1 b)) durchgeführt worden.

b. Hinsichtlich der in der internationalen Anmeldung offenbarten **Nucleotid- und/oder Aminosäuresequenz** ist die internationale Recherche auf der Grundlage des Sequenzprotokolls durchgeführt worden, das

☐ in der internationalen Anmeldung in schriftlicher Form enthalten ist.

☐ zusammen mit der internationalen Anmeldung in computerlesbarer Form eingereicht worden ist.

☐ bei der Behörde nachträglich in schriftlicher Form eingereicht worden ist.

☐ bei der Behörde nachträglich in computerlesbarer Form eingereicht worden ist.

☐ Die Erklärung, daß das nachträglich eingereichte schriftliche Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht, wurde vorgelegt.

☐ Die Erklärung, daß die in computerlesbarer Form erfaßten Informationen dem schriftlichen Sequenzprotokoll entsprechen, wurde vorgelegt.

2. ☐ Bestimmte Ansprüche haben sich als nicht recherchierbar erwiesen (siehe Feld I).

3. ☐ Mangelnde Einheitlichkeit der Erfindung (siehe Feld II).

4. Hinsichtlich der Bezeichnung der Erfindung

☒ wird der vom Anmelder eingereichte Wortlaut genehmigt.

☐ wurde der Wortlaut von der Behörde wie folgt festgesetzt:

5. Hinsichtlich der Zusammenfassung

☒ wird der vom Anmelder eingereichte Wortlaut genehmigt.

☐ wurde der Wortlaut nach Regel 38.2b) in der in Feld III angegebenen Fassung von der Behörde festgesetzt. Der Anmelder kann der Behörde innerhalb eines Monats nach dem Datum der Absendung dieses internationalen Recherchenberichts eine Stellungnahme vorlegen.

6. Folgende Abbildung der Zeichnungen ist mit der Zusammenfassung zu veröffentlichen: Abb. Nr.

☐ wie vom Anmelder vorgeschlagen

☐ weil der Anmelder selbst keine Abbildung vorgeschlagen hat.

☐ weil diese Abbildung die Erfindung besser kennzeichnet.

☒ keine der Abb.

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 6 G10L5/04

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 5 528 725 A (HUI SIEW K) 18. Juni 1996 (1996-06-18) Spalte 5, Zeile 9 - Zeile 24; Abbildung 2 Spalte 6, Zeile 57 - Spalte 7, Zeile 47; Abbildung 5	1-4, 6, 7, 10
Y	---	8, 9
Y	EP 0 519 802 A (SEXTANT AVIONIQUE) 23. Dezember 1992 (1992-12-23) Seite 3, Spalte 3, Zeile 57 - Seite 4, Zeile 26; Abbildungen 3-6	8, 9
A	EP 0 703 565 A (IBM) 27. März 1996 (1996-03-27) Zusammenfassung Seite 4, Zeile 5 - Seite 6, Zeile 34; Abbildungen 2, 3	1, 8-10

	-/--	

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"Z" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

5. Oktober 1999

Absendedatum des internationalen Recherchenberichts

12/10/1999

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Wanzele, R

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>EVANGELISTA G: "PITCH-SYNCHRONOUS WAVELET REPRESENTATIONS OF SPEECH AND MUSIC SIGNALS"</p> <p>IEEE TRANSACTIONS ON SIGNAL PROCESSING, Bd. 41, Nr. 12, 1. Dezember 1993 (1993-12-01), Seiten 3313-3330, XP000426651</p> <p>ISSN: 1053-587X</p> <p>Zusammenfassung</p> <p>Absatz '00II!; Abbildung 1</p> <p>Absatz '000V!</p> <p>Absatz '0V.A!; Abbildung 7</p> <p>-----</p>	1,10

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

T/DE 99/01308

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
US 5528725	A	18-06-1996	GB	2272554 A	18-05-1994
<hr/>					
EP 0519802	A	23-12-1992	FR	2678103 A	24-12-1992
			WO	9222890 A	23-12-1992
			JP	6503186 T	07-04-1994
			US	5826232 A	20-10-1998
<hr/>					
EP 0703565	A	27-03-1996	JP	8095589 A	12-04-1996
			US	5671330 A	23-09-1997
<hr/>					

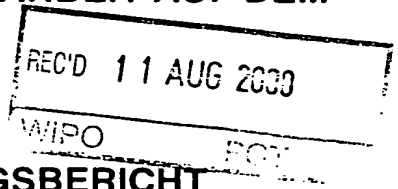
L.T.

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

PCT

INTERNATIONALER VORLÄUFIGER PRÜFUNGSBERICHT

(Artikel 36 und Regel 70 PCT)



Aktenzeichen des Anmelders oder Anwalts 98P1695P	WEITERES VORGEHEN siehe Mitteilung über die Übersendung des internationalen vorläufigen Prüfungsbericht (Formblatt PCT/IPEA/416)	
Internationales Aktenzeichen PCT/DE99/01308	Internationales Anmeldedatum (Tag/Monat/Jahr) 03/05/1999	Prioritätsdatum (Tag/Monat/Tag) 11/05/1998
Internationale Patentklassifikation (IPK) oder nationale Klassifikation und IPK G10L5/04		
Anmelder SIEMENS AKTIENGESELLSCHAFT et al.		

1. Dieser internationale vorläufige Prüfungsbericht wurde von der mit der internationale vorläufigen Prüfung beauftragte Behörde erstellt und wird dem Anmelder gemäß Artikel 36 übermittelt.

2. Dieser BERICHT umfaßt insgesamt 4 Blätter einschließlich dieses Deckblatts.

☒ Außerdem liegen dem Bericht ANLAGEN bei; dabei handelt es sich um Blätter mit Beschreibungen, Ansprüchen und/oder Zeichnungen, die geändert wurden und diesem Bericht zugrunde liegen, und/oder Blätter mit vor dieser Behörde vorgenommenen Berichtigungen (siehe Regel 70.16 und Abschnitt 607 der Verwaltungsrichtlinien zum PCT).

Diese Anlagen umfassen insgesamt 2 Blätter.

3. Dieser Bericht enthält Angaben zu folgenden Punkten:

- I ☒ Grundlage des Berichts
- II ☐ Priorität
- III ☐ Keine Erstellung eines Gutachtens über Neuheit, erfinderische Tätigkeit und gewerbliche Anwendbarkeit
- IV ☐ Mangelnde Einheitlichkeit der Erfindung
- V ☒ Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderische Tätigkeit und der gewerbliche Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung
- VI ☐ Bestimmte angeführte Unterlagen
- VII ☐ Bestimmte Mängel der internationalen Anmeldung
- VIII ☐ Bestimmte Bemerkungen zur internationalen Anmeldung

Datum der Einreichung des Antrags 02/11/1999	Datum der Fertigstellung dieses Berichts 09.08.2000
Name und Postanschrift der mit der internationalen vorläufigen Prüfung beauftragten Behörde: <div style="display: flex; align-items: center;"> <div> Europäisches Patentamt D-80298 München Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465 </div> </div>	Bevollmächtigter Bediensteter Greiser, N Tel. Nr. +49 89 2399 7402



INTERNATIONALER VORLÄUFIGER PRÜFUNGSBERICHT

Internationales Aktenzeichen PCT/DE99/01308

I. Grundlage des Berichts

1. Dieser Bericht wurde erstellt auf der Grundlage (*Ersatzblätter, die dem Anmeldeamt auf eine Aufforderung nach Artikel 14 hin vorgelegt wurden, gelten im Rahmen dieses Berichts als "ursprünglich eingereicht" und sind ihm nicht beigelegt, weil sie keine Änderungen enthalten.*):

Beschreibung, Seiten:

2-11 ursprüngliche Fassung

1,1a eingegangen am 07/07/2000 mit Schreiben vom 20/06/2000

Patentansprüche, Nr.:

1-10 ursprüngliche Fassung

Zeichnungen, Blätter:

1/3-3/3 ursprüngliche Fassung

2. Aufgrund der Änderungen sind folgende Unterlagen fortgefallen:

- ☐ Beschreibung, Seiten:
- ☐ Ansprüche, Nr.:
- ☐ Zeichnungen, Blatt:

3. ☐ Dieser Bericht ist ohne Berücksichtigung (von einigen) der Änderungen erstellt worden, da diese aus den angegebenen Gründen nach Auffassung der Behörde über den Offenbarungsgehalt in der ursprünglich eingereichten Fassung hinausgehen (Regel 70.2(c)):

4. Etwaige zusätzliche Bemerkungen:

V. Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderischen Tätigkeit und der gewerblichen Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung

1. Feststellung

Neuheit (N)	Ja: Ansprüche	1-10
	Nein: Ansprüche	
Erfinderische Tätigkeit (ET)	Ja: Ansprüche	1-10
	Nein: Ansprüche	
Gewerbliche Anwendbarkeit (GA)	Ja: Ansprüche	1-10
	Nein: Ansprüche	

2. Unterlagen und Erklärungen

siehe Beiblatt

zu Punkt V:

1. Die Erfindung betrifft ein Verfahren zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung.

Das Dokument D1= US-A-5528725 (=nächster Stand der Technik) offenbart ein Verfahren zur Bestimmung spektraler Sprachcharakteristika, bei dem eine digitalisierte Äußerung einer Wavelet-Transformation unterzogen wird; das Verfahren wird zur automatischen Spracherkennung verwendet.

Der Erfinder erkennt die technische Aufgabe, daß das in D1 offenbarte Verfahren spektrale Sprachcharakteristika nicht in der Weise bestimmt, um mit dessen Hilfe eine natürlich wirkende synthetische Sprachausgabe zu ermöglichen. Zur Lösung der technischen Aufgabe schlägt deshalb der Erfinder vor, sprecherspezifische Charakteristika in den unterschiedlichen Transformationsstufen der Wavelet-Transformation zu bestimmen.

Der Gegenstand des Anspruchs ist neu. Darüber hinaus gibt es im Stand der Technik keinen Hinweis, der den Fachmann zum Gegenstand des Anspruchs 1 führen würde, sodaß der Anspruch nicht naheliegend ist und folglich einen erfinderischen Schritt beinhaltet.

2. Der Anspruch 10 beansprucht eine Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung. Da der Vorrichtungsanspruch 10 dem Verfahrensanspruch 1 entspricht, ist Anspruch 10 ebenfalls neu und erfinderisch.

M 07.00

1

Beschreibung**Verfahren und Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung**

5

Die Erfindung betrifft ein Verfahren und eine Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung.

- 10 Bei einer konkatenativen Sprachsynthese werden einzelne Laute aus Sprachdatenbanken zusammengesetzt. Um dabei einen für das menschliche Ohr natürlich klingenden Sprachverlauf zu erhalten, sind Diskontinuitäten an den Punkten, wo die Laute zusammengesetzt werden (Konkatenationspunkte) zu vermeiden.
- 15 Die Laute sind dabei insbesondere Phoneme einer Sprache oder eine Zusammensetzung mehrerer Phoneme.

- Eine Wavelet-Transformation ist aus [1] bekannt. Bei der Wavelet-Transformation ist durch ein Wavelet-Filter
- 20 gewährleistet, daß jeweils ein Hochpaßanteil und ein Tiefpaßanteil einer nachfolgenden Transformationsstufe ein Signal einer aktuellen Transformationsstufe vollständig wiederherstellen. Dabei erfolgt von einer Transformationsstufe zur nächsten eine Reduktion der
- 25 Auflösung des Hochpaßanteils bzw. Tiefpaßanteils (engl. Fachbegriff: "Subsampling"). Insbesondere ist durch das Subsampling die Anzahl der Transformationsstufen endlich.

- US-A-5528725 offenbart ein Verfahren zur Spracherkennung
- 30 mittels Wavelet-Transformationen.

- EP-A-0519802 offenbart ein Verfahren zur Sprachsynthese, das sprecherspezifische Charakteristika im Hinblick auf eine natürlich klingende Aneinanderreihung von Sprachlauten
- 35 anpasst.

M 07.07.00

1a

Die **Aufgabe** der Erfindung besteht darin, ein Verfahren und eine Anordnung zur Bestimmung spektraler Sprachcharakteristika anzugeben, mit deren Hilfe insbesondere eine natürlich wirkende synthetische Sprachausgabe bestimmbar

5 ist. -

Diese Aufgabe wird gemäß den Merkmalen der unabhängigen Patentansprüche gelöst.

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

Absender: INTERNATIONALE RECHERCHENBEHÖRDE

PCT

An

SIEMENS AKTIENGESELLSCHAFT
Postfach 22 17 34
D-80506 München
GERMANY

ZT GG VM Mon

Eing. 15. Okt. 1999

GR
Frist

MITTEILUNG ÜBER DIE ÜBERMITTLUNG DES
INTERNATIONALEN RECHERCHENBERICHTS
ODER DER ERKLÄRUNG

(Regel 44.1 PCT)

Absendedatum
(Tag/Monat/Jahr)

12/10/1999

Aktenzeichen des Anmelders oder Anwalts

98P1695P

WEITERES VORGEHEN

siehe Punkte 1 und 4 unten

Internationales Aktenzeichen

PCT/DE 99/01308

Internationales Anmeldedatum

(Tag/Monat/Jahr)

03/05/1999

Anmelder

SIEMENS AKTIENGESELLSCHAFT et al.

1. ☒ Dem Anmelder wird mitgeteilt, daß der internationale Recherchenbericht erstellt wurde und ihm hiermit übermittelt wird.

Einreichung von Änderungen und einer Erklärung nach Artikel 19:

Der Anmelder kann auf eigenen Wunsch die Ansprüche der internationalen Anmeldung ändern (siehe Regel 46):

Bis wann sind Änderungen einzureichen?

Die Frist zur Einreichung solcher Änderungen beträgt üblicherweise zwei Monate ab der Übermittlung des internationalen Recherchenberichts; weitere Einzelheiten sind den Anmerkungen auf dem Beiblatt zu entnehmen.

Wo sind Änderungen einzureichen?

Unmittelbar beim Internationalen Büro der WIPO, 34, CHEMIN des Colombettes, CH-1211 Genf 20.
Telefaxnr.: (41-22) 740.14.35

Nähere Hinweise sind den Anmerkungen auf dem Beiblatt zu entnehmen.

2. ☐ Dem Anmelder wird mitgeteilt, daß kein internationaler Recherchenbericht erstellt wird und daß ihm hiermit die Erklärung nach Artikel 17(2)a übermittelt wird.
3. ☐ **Hinsichtlich des Widerspruchs** gegen die Entrichtung einer zusätzlichen Gebühr (zusätzlicher Gebühren) nach Regel 40.2 wird dem Anmelder mitgeteilt, daß
- ☐ der Widerspruch und die Entscheidung hierüber zusammen mit seinem Antrag auf Übermittlung des Wortlauts sowohl des Widerspruchs als auch der Entscheidung hierüber an die Bestimmungsämter dem Internationalen Büro übermittelt worden sind.
- ☐ noch keine Entscheidung über den Widerspruch vorliegt; der Anmelder wird benachrichtigt, sobald eine Entscheidung getroffen wurde.

4. **Weiteres Vorgehen:** Der Anmelder wird auf folgendes aufmerksam gemacht:

Kurz nach Ablauf von **18 Monaten** seit dem Prioritätsdatum wird die internationale Anmeldung vom Internationalen Büro veröffentlicht. Will der Anmelder die Veröffentlichung verhindern oder auf einen späteren Zeitpunkt verschieben, so muß gemäß Regel 90 bis bzw. 90bis vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung eine Erklärung über die Zurücknahme der internationalen Anmeldung oder des Prioritätsanspruchs beim Internationalen Büro eingehen.

Innerhalb von **19 Monaten** seit dem Prioritätsdatum ist ein Antrag auf internationale vorläufige Prüfung einzureichen, wenn der Anmelder den Eintritt in die nationale Phase bis zu 30 Monaten seit dem Prioritätsdatum (in manchen Ämtern sogar noch länger) verschieben möchte.

Innerhalb von **20 Monaten** seit dem Prioritätsdatum muß der Anmelder die für den Eintritt in die nationale Phase vorgeschriebenen Handlungen vor allen Bestimmungsämtern vornehmen, die nicht innerhalb von 19 Monaten seit dem Prioritätsdatum in der Anmeldung oder einer nachträglichen Auswahlerklärung ausgewählt wurden oder nicht ausgewählt werden konnten, da für sie Kapitel II des Vertrages nicht verbindlich ist.

Name und Postanschrift der Internationalen Recherchenbehörde



Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Ahmed Soliman

Diese Anmerkungen sollen grundlegende Hinweise zur Einreichung von Änderungen gemäß Artikel 19 geben. Diesen Anmerkungen liegen die Erfordernisse des Vertrags über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT), der Ausführungsordnung und der Verwaltungsrichtlinien zu diesem Vertrag zugrunde. Bei Abweichungen zwischen diesen Anmerkungen und obengenannten Texten sind letztere maßgebend. Nähere Einzelheiten sind dem PCT-Leitfaden für Anmelder, einer Veröffentlichung der WIPO, zu entnehmen. Die in diesen Anmerkungen verwendeten Begriffe "Artikel", "Regel" und "Abschnitt" beziehen sich jeweils auf die Bestimmungen des PCT-Vertrags, der PCT-Ausführungsordnung bzw. der PCT-Verwaltungsrichtlinien.

HINWEISE ZU ÄNDERUNGEN GEMÄSS ARTIKEL 19

Nach Erhalt des internationalen Recherchenberichts hat der Anmelder die Möglichkeit, einmal die Ansprüche der internationalen Anmeldung zu ändern. Es ist jedoch zu betonen, daß, da alle Teile der internationalen Anmeldung (Ansprüche, Beschreibung und Zeichnungen) während des internationalen vorläufigen Prüfungsverfahrens geändert werden können, normalerweise keine Notwendigkeit besteht, Änderungen der Ansprüche nach Artikel 19 einzureichen, außer wenn der Anmelder z.B. zum Zwecke eines vorläufigen Schutzes die Veröffentlichung dieser Ansprüche wünscht oder ein anderer Grund für eine Änderung der Ansprüche vor ihrer internationalen Veröffentlichung vorliegt. Weiterhin ist zu beachten, daß ein vorläufiger Schutz nur in einigen Staaten erhältlich ist.

Welche Teile der internationalen Anmeldung können geändert werden?

Im Rahmen von Artikel 19 können nur die Ansprüche geändert werden.

In der internationalen Phase können die Ansprüche auch nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert (oder nochmals geändert) werden. Die Beschreibung und die Zeichnungen können nur nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert werden.

Beim Eintritt in die nationale Phase können alle Teile der internationalen Anmeldung nach Artikel 28 oder gegebenenfalls Artikel 41 geändert werden.

Bis wann sind Änderungen einzureichen?

Innerhalb von zwei Monaten ab der Übermittlung des internationalen Recherchenberichts oder innerhalb von sechzehn Monaten ab dem Prioritätsdatum, je nachdem, welche Frist später abläuft. Die Änderungen gelten jedoch als rechtzeitig eingereicht, wenn sie dem Internationalen Büro nach Ablauf der maßgebenden Frist, aber noch vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung (Regel 46.1) zugehen.

Wo sind die Änderungen nicht einzureichen?

Die Änderungen können nur beim Internationalen Büro, nicht aber beim Anmeldeamt oder der Internationalen Recherchenbehörde eingereicht werden (Regel 46.2).

Falls ein Antrag auf internationale vorläufige Prüfung eingereicht wurde/wird, siehe unten.

In welcher Form können Änderungen erfolgen?

Eine Änderung kann erfolgen durch Streichung eines oder mehrerer ganzer Ansprüche, durch Hinzufügung eines oder mehrerer neuer Ansprüche oder durch Änderung des Wortlauts eines oder mehrerer Ansprüche in der eingereichten Fassung.

Für jedes Anspruchsblatt, das sich aufgrund einer oder mehrerer Änderungen von dem ursprünglich eingereichten Blatt unterscheidet, ist ein Ersatzblatt einzureichen.

Alle Ansprüche, die auf einem Ersatzblatt erscheinen, sind mit arabischen Ziffern zu numerieren. Wird ein Anspruch gestrichen, so brauchen, die anderen Ansprüche nicht neu numeriert zu werden. Im Fall einer Neunummerierung sind die Ansprüche fortlaufend zu numerieren (Verwaltungsrichtlinien, Abschnitt 205 b)).

Die Änderungen sind in der Sprache abzufassen, in der die internationale Anmeldung veröffentlicht wird.

Welche Unterlagen sind den Änderungen beizufügen?

Begleitschreiben (Abschnitt 205 b)):

Die Änderungen sind mit einem Begleitschreiben einzureichen.

Das Begleitschreiben wird nicht zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht. Es ist nicht zu verwechseln mit der "Erklärung nach Artikel 19(1)" (siehe unten, "Erklärung nach Artikel 19 (1)").

Das Begleitschreiben ist nach Wahl des Anmelders in englischer oder französischer Sprache abzufassen. Bei englischsprachigen internationalen Anmeldungen ist das Begleitschreiben aber ebenfalls in englischer, bei französischsprachigen internationalen Anmeldungen in französischer Sprache abzufassen.

ANMERKUNGEN ZU FORMBLATT PCT/ISA/220 (Fortsetzung)

Im Begleitschreiben sind die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen anzugeben. So ist insbesondere zu jedem Anspruch in der internationalen Anmeldung anzugeben (gleichlautende Angaben zu verschiedenen Ansprüchen können zusammengefaßt werden), ob

- i) der Anspruch unverändert ist;
- ii) der Anspruch gestrichen worden ist;
- iii) der Anspruch neu ist;
- iv) der Anspruch einen oder mehrere Ansprüche in der eingereichten Fassung ersetzt;
- v) der Anspruch auf die Teilung eines Anspruchs in der eingereichten Fassung zurückzuführen ist.

Im folgenden sind Beispiele angegeben, wie Änderungen im Begleitschreiben zu erläutern sind:

1. [Wenn anstelle von ursprünglich 48 Ansprüchen nach der Änderung einiger Ansprüche 51 Ansprüche existieren]:
"Die Ansprüche 1 bis 29, 31, 32, 34, 35, 37 bis 48 werden durch geänderte Ansprüche gleicher Numerierung ersetzt; Ansprüche 30, 33 und 36 unverändert; neue Ansprüche 49 bis 51 hinzugefügt."
2. [Wenn anstelle von ursprünglich 15 Ansprüchen nach der Änderung aller Ansprüche 11 Ansprüche existieren]:
"Geänderte Ansprüche 1 bis 11 treten an die Stelle der Ansprüche 1 bis 15."
3. [Wenn ursprünglich 14 Ansprüche existierten und die Änderungen darin bestehen, daß einige Ansprüche gestrichen werden und neue Ansprüche hinzugefügt werden]:
"Ansprüche 1 bis 6 und 14 unverändert; Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt. "Oder" Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt; alle übrigen Ansprüche unverändert."
4. [Wenn verschiedene Arten von Änderungen durchgeführt werden]:
"Ansprüche 1-10 unverändert; Ansprüche 11 bis 13, 18 und 19 gestrichen; Ansprüche 14, 15 und 16 durch geänderten Anspruch 14 ersetzt; Anspruch 17 in geänderte Ansprüche 15, 16 und 17 unterteilt; neue Ansprüche 20 und 21 hinzugefügt."

"Erklärung nach Artikel 19(1)" (Regel 46.4)

Den Änderungen kann eine Erklärung beigelegt werden, mit der die Änderungen erläutert und ihre Auswirkungen auf die Beschreibung und die Zeichnungen dargelegt werden (die nicht nach Artikel 19 (1) geändert werden können).

Die Erklärung wird zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht.

Sie ist in der Sprache abzufassen, in der die internationale Anmeldung veröffentlicht wird.

Sie muß kurz gehalten sein und darf, wenn in englischer Sprache abgefaßt oder ins Englische übersetzt, nicht mehr als 500 Wörter umfassen.

Die Erklärung ist nicht zu verwechseln mit dem Begleitschreiben, das auf die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen hinweist, und ersetzt letzteres nicht. Sie ist auf einem gesonderten Blatt einzureichen und in der Überschrift als solche zu kennzeichnen, vorzugsweise mit den Worten "Erklärung nach Artikel 19 (1)".

Die Erklärung darf keine herabsetzenden Äußerungen über den internationalen Recherchenbericht oder die Bedeutung von in dem Bericht angeführten Veröffentlichungen enthalten. Sie darf auf im internationalen Recherchenbericht angeführte Veröffentlichungen, die sich auf einen bestimmten Anspruch beziehen, nur im Zusammenhang mit einer Änderung dieses Anspruchs Bezug nehmen.

Auswirkungen eines bereits gestellten Antrags auf internationale vorläufige Prüfung

Ist zum Zeitpunkt der Einreichung von Änderungen nach Artikel 19 bereits ein Antrag auf internationale vorläufige Prüfung gestellt worden, so sollte der Anmelder in seinem Interesse gleichzeitig mit der Einreichung der Änderungen beim Internationalen Büro auch eine Kopie der Änderungen bei der mit der internationalen vorläufigen Prüfung beauftragten Behörde einreichen (siehe Regel 62.2 a), erster Satz).

Auswirkungen von Änderungen hinsichtlich der Übersetzung der internationalen Anmeldung beim Eintritt in die nationale Phase

Der Anmelder wird darauf hingewiesen, daß bei Eintritt in die nationale Phase möglicherweise anstatt oder zusätzlich zu der Übersetzung der Ansprüche in der eingereichten Fassung eine Übersetzung der nach Artikel 19 geänderten Ansprüche an die bestimmten/ausgewählten Ämter zu übermitteln ist.

Nähere Einzelheiten über die Erfordernisse jedes bestimmten/ausgewählten Amtes sind Band II des PCT-Leitfadens für Anmelder zu entnehmen.

REPLACED BY
ART 34 AMST**METHOD AND ARRANGEMENT FOR DETERMINING SPECTRAL
SPEECH CHARACTERISTICS IN A SPOKEN EXPRESSION**

The invention is directed to a method and to an arrangement for determining spectral speech characteristics in a spoken expression.

5 In a concatenative speech synthesis, individual sounds are combined from speech data banks. In order to thereby obtain a speech curve that sounds natural to the human ear, discontinuities must be avoided at the points where the sounds are combined (concatenation points). In particular, the sounds are thereby phonemes of a language or a combination of a plurality of phonemes.

10 [1] discloses a wavelet transformation. In wavelet transformation, a wavelet filter assures that a respective high-pass part and low-pass part of a following transformation stage completely restore a signal of a current transformation stage. A reduction of the resolution of the high-pass part or, respectively, low-pass part thereby ensues from one transformation stage to the next (English art term: "sub-sampling").

15 In particular, the plurality of transformation stages is finite due to the sub-sampling.

The object of the invention is comprised in specifying a method and an arrangement for determining spectral speech characteristics with whose assistance, in particular, a speech output that sounds natural can be determined.

20 This object is achieved according to the features of the independent claims.

A method for determining spectral speech characteristics in a spoken expression is recited in the scope of the invention. To that end, the spoken expression is digitalized and subjected to a wavelet transformation. The speaker-specific characteristics are determined on the basis of different transformation stages of the
25 wavelet transformation.

Beschreibung**Verfahren und Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung**

5

Die Erfindung betrifft ein Verfahren und eine Anordnung zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung.

- 10 Bei einer konkatenativen Sprachsynthese werden einzelne Laute aus Sprachdatenbanken zusammengesetzt. Um dabei einen für das menschliche Ohr natürlich klingenden Sprachverlauf zu erhalten, sind Diskontinuitäten an den Punkten, wo die Laute zusammengesetzt werden (Konkatenationspunkte) zu vermeiden.
- 15 Die Laute sind dabei insbesondere Phoneme einer Sprache oder eine Zusammensetzung mehrerer Phoneme.

- Eine Wavelet-Transformation ist aus [1] bekannt. Bei der Wavelet-Transformation ist durch ein Wavelet-Filter
- 20 gewährleistet, daß jeweils ein Hochpaßanteil und ein Tiefpaßanteil einer nachfolgenden Transformationsstufe ein Signal einer aktuellen Transformationsstufe vollständig wiederherstellen. Dabei erfolgt von einer Transformationsstufe zur nächsten eine Reduktion der
- 25 Auflösung des Hochpaßanteils bzw. Tiefpaßanteils (engl. Fachbegriff: "Subsampling"). Insbesondere ist durch das Subsampling die Anzahl der Transformationsstufen endlich.

- Die **Aufgabe** der Erfindung besteht darin, ein Verfahren und
- 30 eine Anordnung zur Bestimmung spektraler Sprachcharakteristika anzugeben, mit deren Hilfe insbesondere eine natürlich wirkende synthetische Sprachausgabe bestimmbar ist.

- 35 Diese Aufgabe wird gemäß den Merkmalen der unabhängigen Patentansprüche gelöst.

Im Rahmen der Erfindung wird ein Verfahren angegeben zur Bestimmung spektraler Sprachcharakteristika in einer gesprochenen Äußerung. Dazu wird die gesprochene Äußerung digitalisiert und einer Wavelet-Transformation unterzogen.

5 Anhand unterschiedlicher Transformationsstufen der Wavelet-Transformation werden die sprecherspezifischen Charakteristika ermittelt.

10 Dabei ist es insbesondere ein Vorteil, daß bei der Wavelet-Transformation mittels eines Hochpaßfilters und eines Tiefpaßfilters die Äußerung aufgeteilt wird und unterschiedliche Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen sprecherspezifische Charakteristika enthalten.

15 Die einzelnen Hochpaßanteile bzw. Tiefpaßanteile verschiedener Transformationsstufen stehen für vorgegebene sprecherspezifische Charakteristika, wobei sowohl Hochpaßanteil als auch Tiefpaßanteil einer jeweiligen Transformationsstufe, also das jeweilige Charakteristikum, 20 getrennt von anderen Charakteristika modifiziert werden kann. Setzt man bei der inversen Wavelet-Transformation aus den jeweiligen Hochpaß- und Tiefpaßanteilen der einzelnen Transformationsstufen wieder das ursprüngliche Signal zusammen, so ist gewährleistet, daß genau das gewünschte 25 Charakteristikum verändert worden ist. Es ist somit möglich bestimmte vorgegebene Eigenarten der Äußerung zu verändern, ohne daß dadurch der Rest der Äußerung beeinflusst wird.

30 Eine Ausgestaltung besteht darin, daß vor der Wavelet-Transformation die Äußerung gefenstert, also eine vorgegebene Menge von Abtastwerten ausgeschnitten, und in den Frequenzbereich transformiert wird. Hierzu wird insbesondere eine Fast-Fourier-Transformation (FFT) angewandt.

35 Eine weitere Ausgestaltung besteht darin, daß ein Hochpaßanteil einer Transformationsstufe in einen Realteil

und einen Imaginärteil aufgeteilt wird. Der Hochpaßanteil der Wavelet-Transformation entspricht dem Differenzsignal zwischen dem aktuellen Tiefpaßanteil und dem Tiefpaßanteil der vorhergehenden Transformationsstufe.

5

Insbesondere besteht eine Weiterbildung darin, die Zahl der durchzuführenden Transformationsstufen der Wavelet-Transformation dadurch zu bestimmen, daß in der letzten Transformationsstufe, die aus hintereinandergeschalteten

10 Tiefpässen besteht, ein Gleichanteil der Äußerung enthalten ist. Dann ist das Signal als Ganzes darstellbar durch seine Wavelet-Koeffizienten. Dies entspricht der vollständigen Transformation der Information des Signalausschnitts in den Wavelet-Raum.

15

Wird insbesondere nur der jeweilige Tiefpaßanteil weiter transformiert (mittels eines Hochpaß- und eines Tiefpaßfilters), so verbleibt als Hochpaßanteil einer Transformationsstufe das Differenzsignal, wie oben erläutert.

20 Kumuliert man Differenzsignale (Hochpaßanteile) über die Transformationsstufen, erhält man in der letzten Transformationsstufe als kumulierten Hochpaßanteil die Information der gesprochenen Äußerung ohne Gleichanteil.

25

Im Rahmen einer zusätzlichen Weiterbildung sind die sprecherspezifischen Charakteristika identifizierbar als:

a) Grundfrequenz:

30

Die Schwingung des Hochpaßanteils der ersten oder der zweiten Transformationsstufe der Wavelet-Transformation läßt die Grundfrequenz der Äußerung erkennen. Die Grundfrequenz zeigt an, ob der Sprecher ein Mann oder eine Frau ist.

35

b) Form der spektralen Hüllkurve:

Die spektrale Hüllkurve enthält Information über eine Transferfunktion des Vokaltrakts bei der Artikulation.

In einem stimmhaften Bereich wird die spektrale Hüllkurve von den Formanten dominiert. Der Hochpaßanteil einer höheren Transformationsstufe der Wavelet-Transformation enthält diese spektrale Hüllkurve.

c) Spectral Tilt (Rauchigkeit):

Die Rauchigkeit in einer Stimme wird als negative Steigung im Verlauf des vorletzten Tiefpaßanteils sichtbar.

Die sprecherspezifischen Charakteristika a) bis c) sind bei der Sprachsynthese von großer Bedeutung. Wie eingangs erwähnt, bedient man sich bei der konkatenativen Sprachsynthese großer Mengen realgesprochener Äußerungen, aus denen Beispiellaute ausgeschnitten und später zu einem neuen Wort zusammengesetzt werden (synthetisierte Sprache). Dabei sind Diskontinuitäten zwischen zusammengesetzten Lauten von Nachteil, da diese vom menschlichen Ohr als unnatürlich wahrgenommen werden. Um den Diskontinuitäten entgegenzuwirken ist es von Vorteil, direkt die perzeptiv relevanten Größen zu erfassen und ggf. zu vergliche und/oder einander anzupassen.

Dies kann geschehen durch direkte Manipulation, indem ein Sprachlaut in mindestens einer seiner sprecherspezifischen Charakteristika angepaßt wird, so daß er in dem akustischen Kontext der konkatenativ verknüpften Laute nicht als störend wahrgenommen wird. Auch ist es möglich, die Auswahl eines passenden Lautes daran auszurichten, daß sprecherspezifische Charakteristika von zu verknüpfenden Lauten möglichst gut zueinander passen, z.B. daß den Lauten gleiche oder ähnliche Rauchigkeit zu eigen ist.

Ein Vorteil der Erfindung besteht darin, daß die spektrale Hüllkurve den Artikulationstrakt des Sprechers widerspiegelt und nicht, wie z.B. ein Polstellenmodell, auf Formanten gestützt ist. Weiterhin gehen bei der Wavelet-Transformation

als nichtparametrischer Darstellung keine Daten verloren, die Äußerung kann stets vollständig rekonstruiert werden. Die aus den einzelnen Transformationsstufen der Wavelet-Transformation hervorgehenden Daten sind linear voneinander unabhängig, können somit getrennt voneinander beeinflußt und später wieder zu der beeinflußten Äußerung - verlustlos - zusammengesetzt werden.

- 10 Weiterhin wird eine Anordnung zur Bestimmung spektraler Sprachcharakteristika angegeben, die eine Prozessoreinheit aufweist, die derart eingerichtet ist, daß eine Äußerung digitalisierbar ist. Daraufhin wird die Äußerung einer Wavelet-Transformation unterzogen und anhand
15 unterschiedlicher Transformationsstufen werden sprecherspezifische Charakteristika ermittelt.

Diese Anordnung ist insbesondere geeignet zur Durchführung des erfindungsgemäßen Verfahrens oder einer seiner vorstehend
20 erläuterten Weiterbildungen.

Weiterbildungen der Erfindung ergeben sich auch aus den abhängigen Ansprüchen.

- 25 Ausführungsbeispiele der Erfindung werden nachfolgend anhand der Zeichnung dargestellt und erläutert.

Es zeigen

- 30 Fig.1 eine Wavelet-Funktion;
Fig.2 eine Wavelet-Funktion, unterteilt nach Realteil und Imaginärteil;
35 Fig.3 eine kaskadierte Filterstruktur, die die Transformationsschritte der Wavelet-Transformation darstellt;

Fig.4 Tiefpaßanteile und Hochpaßanteile unterschiedlicher Transformationsstufen;

5 Fig.5 Schritte der konkatentativen Sprachsynthese.

Fig.1 zeigt eine Wavelet-Funktion, die bestimmt ist durch

$$10 \quad \psi(f) = c \cdot \left(1 - \left(\frac{f}{\sigma}\right)^2\right) \cdot e^{-\frac{1}{2} \cdot \left(\frac{f}{\sigma}\right)^2} \quad (1),$$

wobei

f die Frequenz,

σ eine Standardabweichung und

15 c eine vorgegebene Normierungskonstante
bezeichnen.

Insbesondere ist die Standardabweichung σ bestimmt durch die vorgebbare Stelle des Seitenbandminimums 101 in Fig.1.

20

Fig.2 zeigt eine Wavelet-Funktion mit einem Realteil gemäß Gleichung (1) und einer Hilbert-Transformierten H des Realteils als Imaginärteil. Die komplexe Wavelet-Funktion ergibt sich somit zu

25

$$\Psi(f) = \psi(f) + j \cdot H\{\psi(f)\} \quad (2).$$

Die Konstante c aus Gleichung (1) wird verwendet, um die komplexe Wavelet-Funktion zu normieren:

30

$$\int_{-\infty}^{\infty} \Psi(f) \cdot \bar{\Psi}(f) df = 1 \quad (3),$$

wobei $\bar{\Psi}$ die konjugiert komplexe Wavelet-Funktion bezeichnet.

Fig.3 zeigt die kaskadierte Anwendung der Wavelet-Transformation. Ein Signal 301 wird sowohl durch einen Hochpaß HP1 302 als auch durch einen Tiefpaß TP1 305 gefiltert. Dabei findet insbesondere ein Subsampling statt, d.h. die Anzahl der abzuspeichernden Werte wird pro Filter reduziert. Eine inverse Wavelet-Transformation gewährleistet, daß aus dem Tiefpaßanteil TP1 305 und dem Hochpaßanteil HP1 304 wieder das ursprüngliche Signal 301 rekonstruierbar ist.

Im Hochpaß HP1 302 wird getrennt nach Realteil Re1 303 und Imaginärteil Im1 304 gefiltert.

Das Signal 310 nach dem Tiefpaßfilter TP1 305 wird erneut sowohl durch einen Hochpaß HP2 306 als auch durch einen Tiefpaß TP2 309 gefiltert. Der Hochpaß HP2 306 umfaßt wieder einen Realteil Re2 307 und einen Imaginärteil Im2 308. Das Signal nach der zweiten Transformationsstufe 311 wird wieder gefiltert, usf.

Geht man von einem (FFT-transformierten) Kurzzeitspektrum mit 256 Werten aus, so werden acht Transformationsschritte durchgeführt (Subsamplingrate: 1/2), bis das Signal aus dem letzten Tiefpaßfilter TP8 dem Gleichanteil entspricht.

In **Fig.4** sind verschiedene Transformationsstufen der Wavelet-Transformation, unterteilt nach Tiefpaßanteilen (Figuren 4A, 4C und 4E) und Hochpaßanteilen (Figuren 4B, 4D und 4F) dargestellt.

Aus dem Hochpaßanteil gemäß Fig.4B ist die Grundfrequenz der gesprochenen Äußerung ersichtlich. Neben den Schwankungen in der Amplitude ist deutlich eine überwiegende Periodizität im wavelet-gefilterten Spektrum zu erkennen, die Grundfrequenz des Sprechers. Anhand der Grundfrequenz ist es möglich, vorgegebene Äußerungen bei der Sprachsynthese einander

anzupassen oder passende Äußerungen aus einer Datenbank mit vorgegebene Äußerungen zu bestimmen.

Im Tiefpaßanteil von Fig.4C sind als ausgeprägte Minima und Maxima die Formanten des Sprachsignalausschnitts (die Länge des Sprachsignalausschnitts entspricht in etwa der doppelten Grundfrequenz) dargestellt. Die Formanten repräsentieren Resonanzfrequenzen im Vokaltrakt des Sprechers. Die deutliche Darstellbarkeit der Formanten ermöglicht eine Anpassung und/oder Auswahl passender Lautbausteine bei der konkatenativen Sprachsynthese.

Im Tiefpaßanteil der vorletzten Transformationsstufe (bei 256 Frequenzwerten im Originalsignal: TP7), kann die Rauchigkeit einer Stimme ermittelt werden. Der Abstieg des Kurvenverlaufs zwischen Maximum Mx und Minimum Mi kennzeichnet den Grad der Rauchigkeit.

Die erwähnten drei sprecherspezifischen Charakteristika sind somit identifiziert und können für die Sprachsynthese gezielt beeinflußt werden. Dabei ist es insbesondere von Bedeutung, daß bei der inversen Wavelet-Transformation die Manipulation eines einzelnen sprecherspezifischen Charakteristikums nur dieses beeinflußt, die anderen perzeptiv relevanten Größen bleiben unberührt. Somit kann die Grundfrequenz gezielt verstellt werden, ohne daß dadurch die Rauchigkeit der Stimme beeinflußt wird.

Eine andere Einsatzmöglichkeit besteht in der Auswahl eines geeigneten Lautabschnitts zur konkatenativen Verknüpfung mit einem anderen Lautabschnitt, wobei beide Lautabschnitte ursprünglich von verschiedenen Sprechern in unterschiedlichen Kontexten aufgenommen wurden. Mit Ermittlung spektraler Sprachcharakteristika kann ein geeigneter zu verknüpfender Lautabschnitt gefunden werden, da mit den Charakteristika Kriterien bekannt sind, die einen Vergleich von Lautabschnitten untereinander und somit eine Auswahl des

passenden Lautabschnitts automatisch nach bestimmten Vorgaben ermöglichen.

Fig.5 zeigt Schritte einer konkatenativen Sprachsynthese.

- 5 Eine Datenbank wird mit einer vorgegebenen Menge natürlichgesprochener Sprache verschiedener Sprecher erstellt, wobei Lautabschnitte in der natürlichgesprochenen Sprache identifiziert und abgespeichert werden. Es ergeben sich zahlreiche Repräsentanten für die verschiedenen
- 10 Lautabschnitte einer Sprache, auf die die Datenbank zugreifen kann. Die Lautabschnitte sind insbesondere Phoneme einer Sprache oder eine Aneinanderreihung solcher Phoneme. Je kleiner der Lautabschnitt, desto größer sind die Möglichkeiten bei der Zusammensetzung neuer Wörter. So umfaßt
- 15 die deutsche Sprache eine vorgegebene Menge von ca. 40 Phonemen, die zur Synthese nahezu aller Wörter der Sprache ausreichen. Dabei sind unterschiedliche akustische Kontexte zu berücksichtigen, je nachdem, in welchem Wort das jeweilige Phonem auftritt. Nun ist es wichtig, die einzelnen Phoneme in
- 20 den akustischen Kontext derart einzubetten, daß Diskontinuitäten, die vom menschlichen Gehör als unnatürlich und "synthetisch" empfunden werden, vermieden werden. Wie erwähnt stammen die Lautabschnitte von unterschiedlichen Sprechern und weisen somit verschiedene sprecherspezifische
- 25 Charakteristika auf. Um eine möglichst natürlich wirkende Äußerung zu synthetisieren, ist es wichtig, die Diskontinuitäten zu minimieren. Dies kann erfolgen durch Anpassung der identifizierbaren und modifizierbaren sprecherspezifischen Charakteristika oder durch Auswahl
- 30 passender Lautabschnitte aus der Datenbank, wobei ebenfalls die sprecherspezifischen Charakteristika bei der Auswahl ein entscheidendes Hilfsmittel darstellen.

- In Fig.5 sind beispielhaft zwei Laute A 507 und B 508
- 35 dargestellt, die jeweils einzelne Lautabschnitte 505 bzw. 506 aufweisen. Die Laute A 507 und B 508 stammen jeweils aus einer gesprochenen Äußerung, wobei der Laut A 507 deutlich

vom Laut B 508 verschieden ist. Eine Trennlinie 509 zeigt an, wo der Laut A 507 mit dem Laut B 508 verknüpft werden soll. Im vorliegenden Fall sollen die ersten drei Lautabschnitte des Lautes A 507 mit den letzten drei Lautabschnitten des Lautes B 508 konkatenativ verknüpft werden.

Es wird entlang der Trennlinie 509 ein zeitliches Dehnen oder Stauchen (vergleiche Pfeil 503) der aufeinanderfolgenden Lautabschnitte durchgeführt, um den diskontinuierlichen Eindruck am Übergang 509 zu vermindern.

Eine Variante besteht in einem abrupten Übergang der entlang der Trennlinie 509 geteilten Laute. Dabei kommt es jedoch zu den erwähnten Diskontinuitäten, die das menschliche Gehör als störend wahrnimmt. Fügt man hingegen einen Laut C zusammen, daß die Lautabschnitte innerhalb eines Übergangsbereichs 501 oder 502 berücksichtigt werden, wobei ein spektrales Abstandsmaß zwischen zwei einander zuordenbaren Lautabschnitten in dem jeweiligen Übergangsbereich 501 oder 502 angepaßt wird (allmählicher Übergang zwischen den Lautabschnitten). Als das Abstandsmaß herangezogen wird insbesondere im Wavelet-Raum der euklidische Abstand zwischen den in diesem Bereich relevanten Koeffizienten.

Literaturverzeichnis:

- [1] I. Daubechies: "Ten Lectures on Wavelets", Siam Verlag
1992, ISBN 0-89871-274-2, Kapitel 5.1, Seiten 129-137.

Patentansprüche

1. Verfahren zur Bestimmung spektraler Sprachcharakteristika
in einer gesprochenen Äußerung,
 - 5 a) bei dem die Äußerung digitalisiert wird,
 - b) bei dem die digitalisierte Äußerung einer Wavelet-Transformation unterzogen wird,
 - c) bei dem anhand unterschiedlicher Transformationsstufen
der Wavelet-Transformation die sprecherspezifischen
10 Charakteristika bestimmt werden.
2. Verfahren nach Anspruch 1,
bei dem vor der Wavelet-Transformation eine gefensterte
Transformation der digitalisierten Äußerung in einen
15 Frequenzbereich durchgeführt wird.
3. Verfahren nach Anspruch 2,
bei dem die Transformation in den Frequenzbereich mittels
Fast-Fourier-Transformation durchgeführt wird.
20
4. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem in jeder Stufe der Wavelet-Transformation ein
Tiefpaßanteil und ein Hochpaßanteil eines zu
transformierenden Signals ermittelt werden.
25
5. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem ein Hochpaßanteil nach einem Realteil und einem
Imaginärteil unterteilt wird.
- 30 6. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die Wavelet-Transformation mehrere
Transformationsstufen umfaßt, wobei die letzte
Transformationsstufe einen Gleichanteil der Äußerung in
einer der Anzahl Transformationsstufen entsprechenden
35 wiederholten Tiefpaßfilterung liefert.

7. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die sprecherspezifischen Charakteristika bestimmt
sind durch:
- a) eine Grundfrequenz der gesprochenen Äußerung;
 - b) spektrale Hüllkurve;
 - c) einer Rauchigkeit der gesprochenen Äußerung.
8. Verwendung des Verfahrens nach einem der Ansprüche 1 bis
7 zur Sprachsynthese,
wobei einzelne sprecherspezifische Charakteristika im
Hinblick auf eine natürlich klingende Aneinanderreihung
von Sprachlauten angepaßt werden.
9. Verwendung des Verfahrens nach einem der Ansprüche 1 bis
7 zur Sprachsynthese,
wobei aus einer vorgegebenen Datenmenge diejenigen
Sprachlaute anhand einzelner spektraler
Sprachcharakteristika ausgewählt werden, die eine
natürlich klingende Aneinanderreihung von Sprachlauten
gewährleisten.
10. Anordnung zur Bestimmung spektraler Sprachcharakteristika
in einer gesprochenen Äußerung
mit eineressoreinheit, die derart eingerichtet ist,
daß folgende Schritte durchführbar sind:
- a) die Äußerung wird digitalisiert;
 - b) die digitalisierte Äußerung wird einer Wavelet-
Transformation unterzogen;
 - c) anhand unterschiedlicher Transformationsstufen der
Wavelet-Transformation werden die sprecherspezifischen
Charakteristika bestimmt.

Zusammenfassung

Verfahren und Anordnung zur Bestimmung spektraler
Sprachcharakteristika in einer gesprochenen Äußerung

5

Es werden spektrale Sprachcharakteristika in einer
natürlichsprachlichen Äußerung bestimmt, wobei die Äußerung
digitalisiert und einer Wavelet-Transformation unterzogen
wird. Aus den unterschiedlichen Transformationsstufen der

10

Wavelet-Transformation gehen die sprecherspezifischen
Charakteristika hervor. Diese Charakteristika können im
Rahmen einer Sprachsynthese mit Charakteristika anderer
Äußerungen verglichen werden, um ein für das menschliche Ohr
kontinuierlich klingendes synthetisches Sprachsignal zu

15

erzeugen. Alternativ können die Charakteristika auch gezielt
verändert werden, um einer perzeptiven Dissonanz
entgegenzuwirken.

1/3

FIG 1

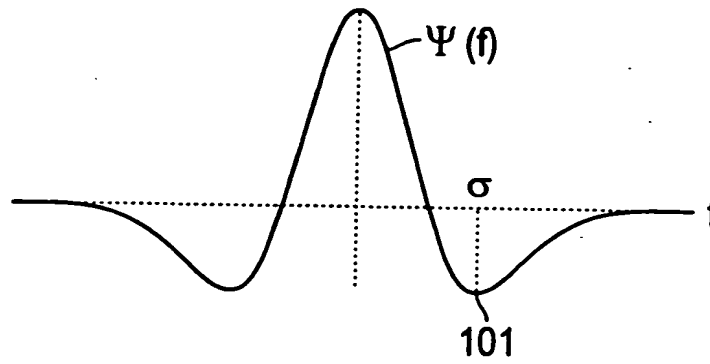


FIG 2

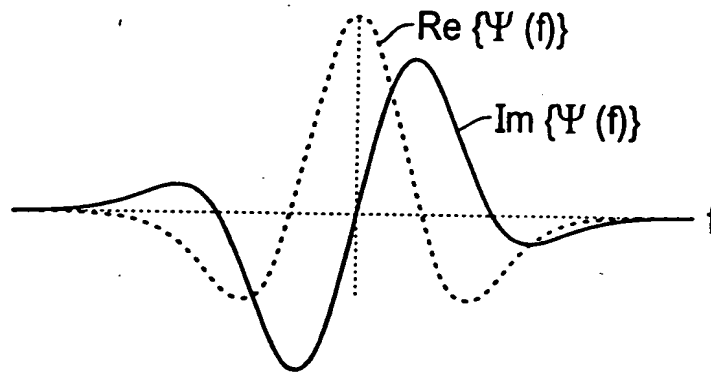
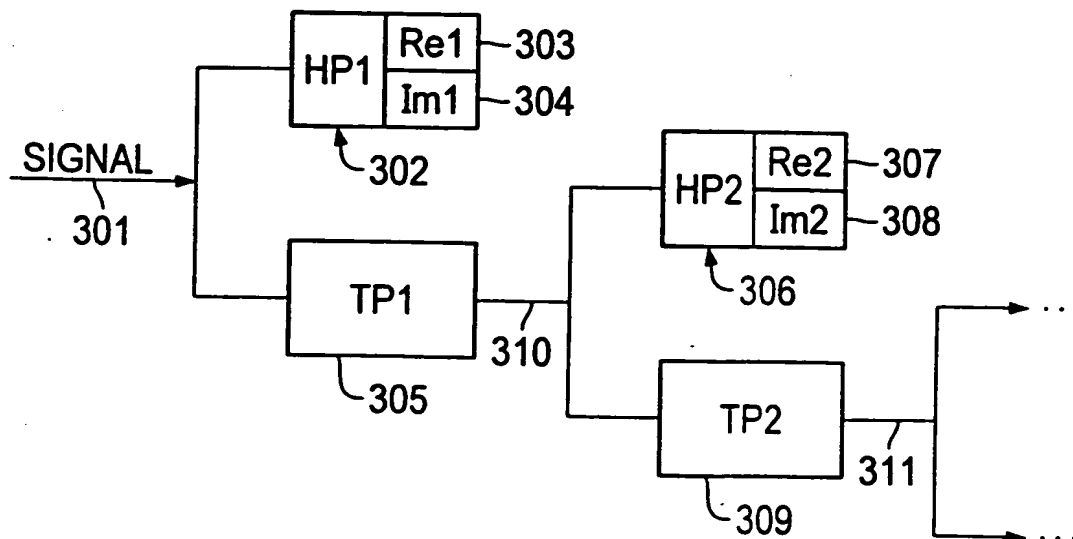


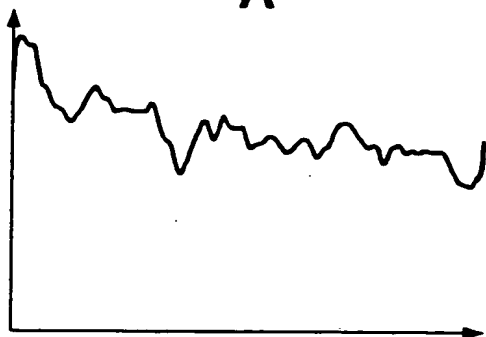
FIG 3



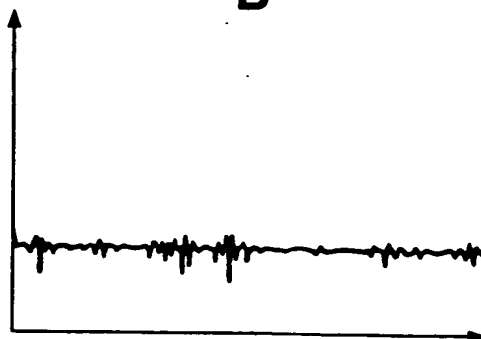
2/3

FIG 4

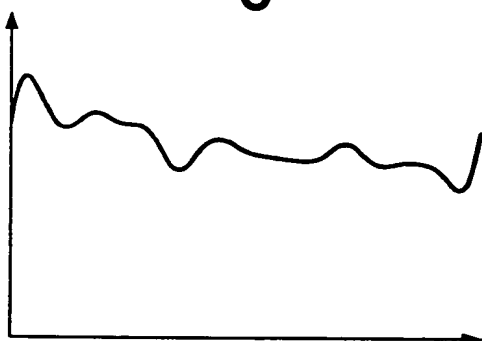
A



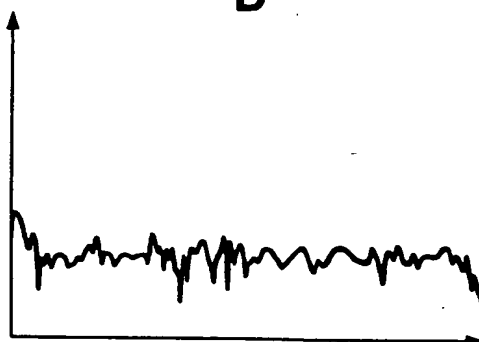
B



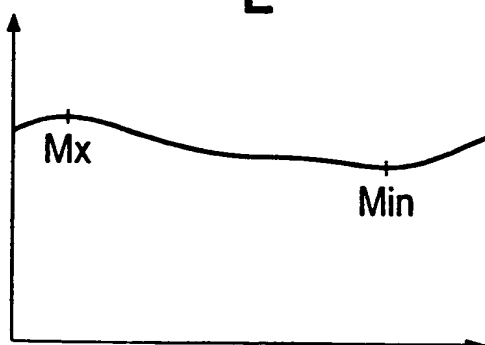
C



D



E



F

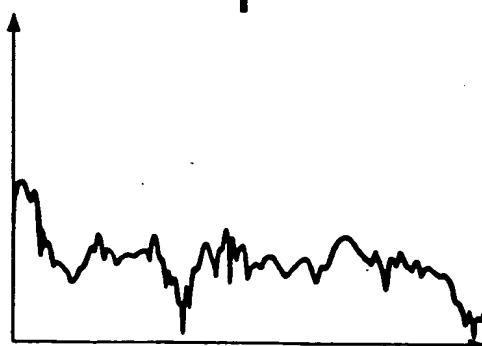
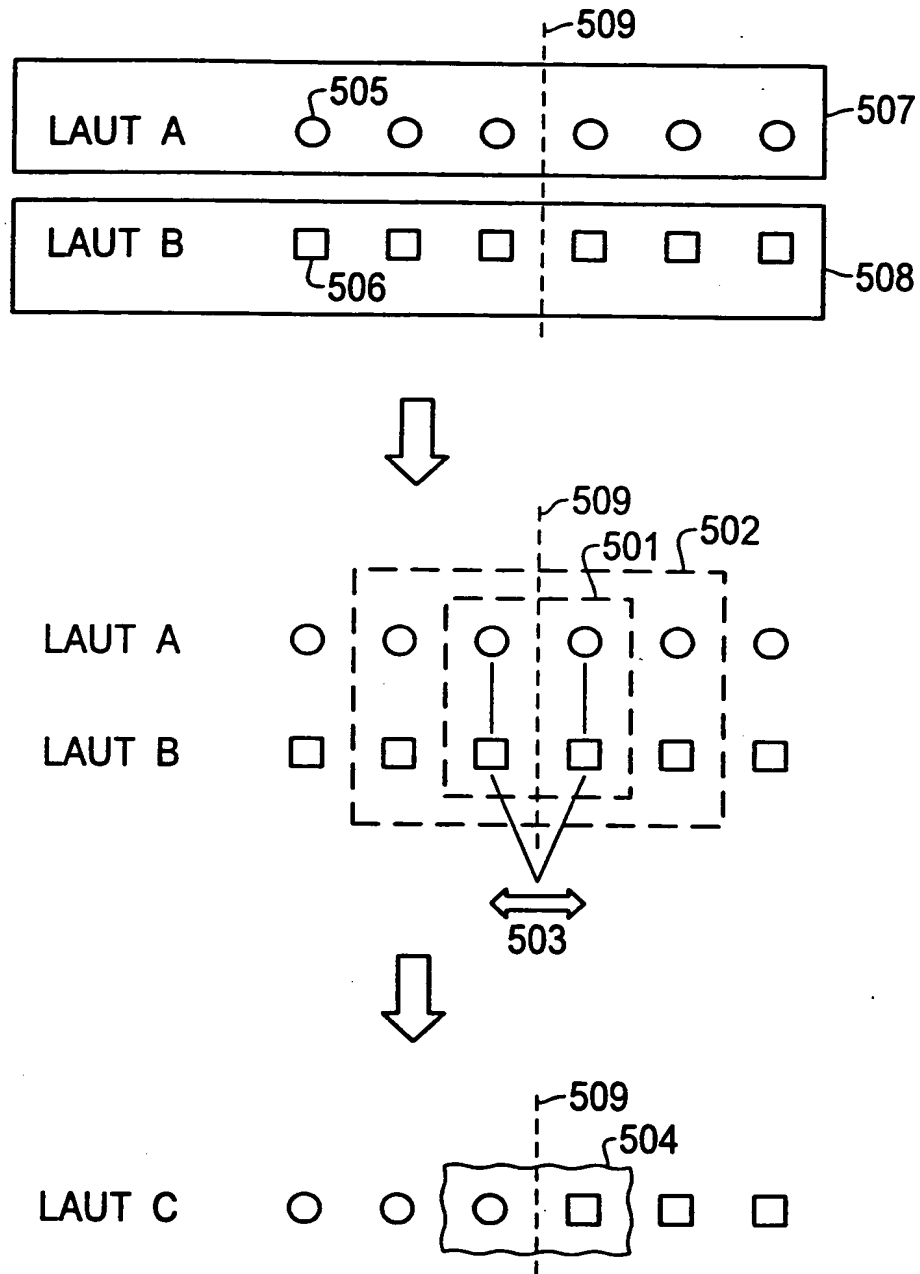


FIG 5



09700311
Translation
2696
-0500

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

2641
3
RECEIVED

JAN 04 2001

Technology Center 2600

Applicant's or agent's file reference 98P1695P	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/DE99/01308	International filing date (day/month/year) 03 May 1999 (03.05.99)	Priority date (day/month/year) 11 May 1998 (11.05.98)
International Patent Classification (IPC) or national classification and IPC G10L 5/04		
Applicant SIEMENS AKTIENGESELLSCHAFT		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets, including this cover sheet.

☒ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 2 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 02 November 1999 (02.11.99)	Date of completion of this report 09 August 2000 (09.08.2000)
Name and mailing address of the IPEA/EP	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/DE99/01308

I. Basis of the report

1. This report has been drawn on the basis of (*Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

- ☐ the international application as originally filed.
- ☒ the description, pages 2-11, as originally filed,
 pages _____, filed with the demand,
 pages 1, 1a, filed with the letter of 20 June 2000 (20.06.2000),
 pages _____, filed with the letter of _____.
- ☒ the claims, Nos. 1-10, as originally filed,
 Nos. _____, as amended under Article 19,
 Nos. _____, filed with the demand,
 Nos. _____, filed with the letter of _____,
 Nos. _____, filed with the letter of _____.
- ☒ the drawings, sheets/fig 1/3-3/3, as originally filed,
 sheets/fig _____, filed with the demand,
 sheets/fig _____, filed with the letter of _____,
 sheets/fig _____, filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages _____
- ☐ the claims, Nos. _____
- ☐ the drawings, sheets/fig _____

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/DE 99/01308

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**1. Statement**

Novelty (N)	Claims	1-10	YES
	Claims		NO
Inventive step (IS)	Claims	1-10	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-10	YES
	Claims		NO

2. Citations and explanations

1. The invention concerns a method for determining spectral voice characteristics in a spoken utterance.

Document D1, US-A-5 528 725, considered the closest prior art, discloses a method for determining spectral voice characteristics in which a digitised utterance is subjected to wavelet transformation. The method is used for automatic voice recognition. The inventors identify the technical problem that the method disclosed in D1 does not determine spectral voice characteristics in such a way as to enable a natural-sounding synthetic speech output. To solve the technical problem, the inventors thus propose determining speaker-specific characteristics in the different transformation steps of the wavelet transformation.

The subject matter of the claim is novel.

Furthermore, the prior art suggests nothing that would lead a person skilled in the art to the subject matter of Claim 1, and therefore the claim is non-obvious and consequently involves an inventive step.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/DE 99/01308

2. Claim 10 concerns a device for determining spectral voice characteristics in a spoken utterance. Since device Claim 10 corresponds to method Claim 1, Claim 10 is likewise novel and inventive.